



COVID-19

FAQ: COVID-19 Data and Surveillance

Frequently Asked Questions

Updated July 23, 2021 [Print](#)

- Below are answers to commonly asked questions about CDC's process for monitoring COVID-19 cases across the country, as well as collecting and sharing data with the public.
- [COVID Data Tracker](#) serves as CDC's home for COVID-19 data.

National COVID-19 Case Surveillance

What is national case surveillance?

Public health departments routinely collect information on people with certain infections. This process, known as case surveillance, can help officials understand where, when, and in which populations an illness is transmitted. This supports action to control outbreaks and prevent the spread of disease. Nationally, more than 120 diseases and conditions are tracked. The information collected is referred to as case data. These efforts are part of a wider ongoing process called public health surveillance. The goal is to protect Americans from infectious diseases and other health threats.

Case surveillance is especially important for new diseases, such as COVID-19. The information collected helps identify similarities and differences among cases. Information commonly collected includes:

- Demographic information (age, race, ethnicity, etc.)
- Clinical factors, such as symptoms
- Epidemiologic characteristics (where, when, and in which populations an illness is transmitted)
- Exposure and contact history (how an illness is spreading)
- Course of clinical illness and care received

State, local, and territorial health departments transmit case data to CDC through the [National Notifiable Diseases Surveillance System](#) (NNDSS). This process is voluntary. To protect individual privacy, all information transmitted to CDC is de-identified. This means it does not contain any personal identifiers, such as names or home addresses, and cannot be linked to an individual. Learn more on [How We Conduct Case Surveillance](#).

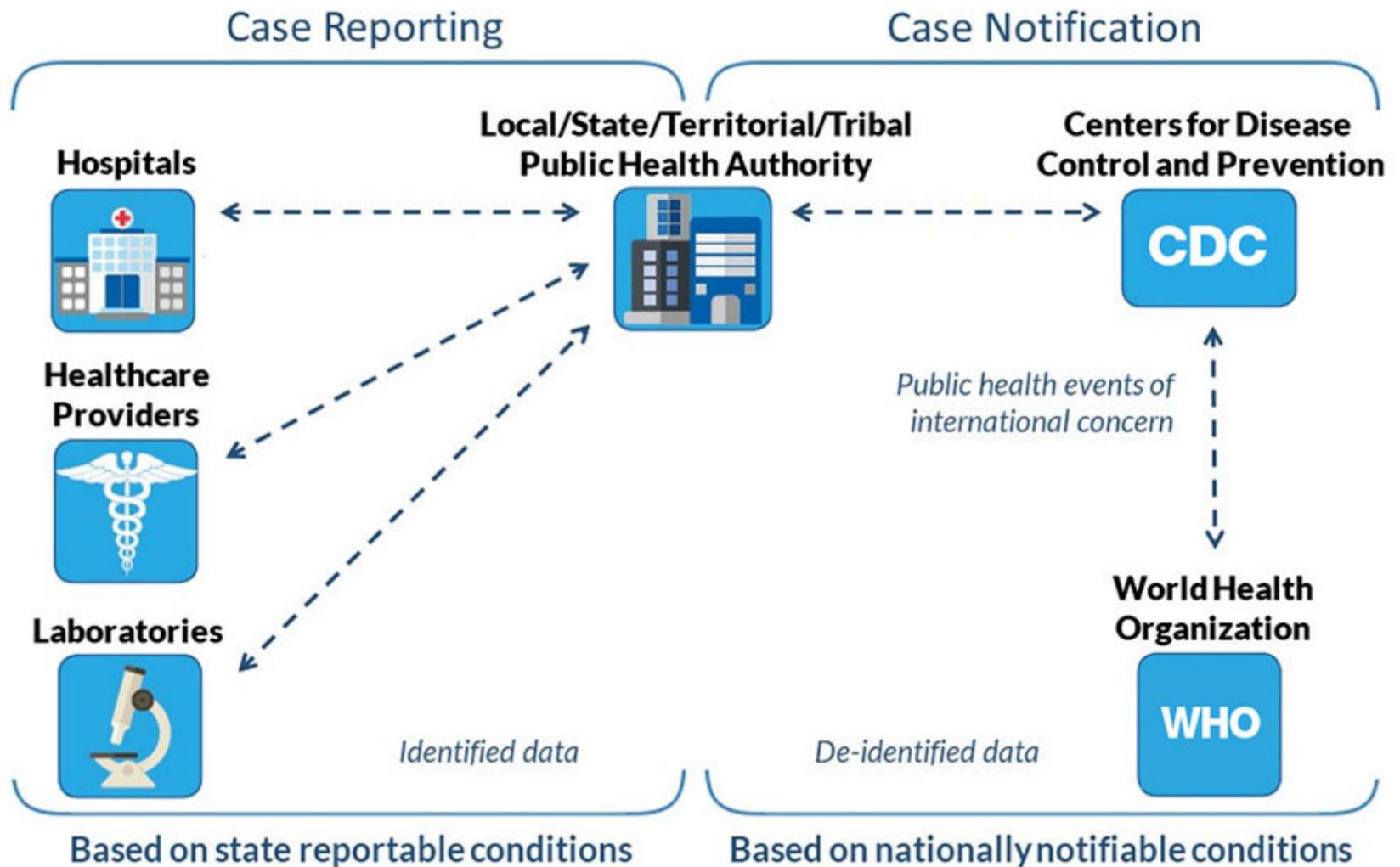
How is COVID-19 case information collected and reported?

Case information is gathered through a data supply chain, which is a process for reporting, collecting, and analyzing disease data. Steps include the following:

- Hospitals, healthcare providers, and laboratories transfer data for case reporting to state, local, and territorial public health departments as required under state disease reporting laws.

public health departments as required under state disease reporting laws.

Case Surveillance



[View Larger](#)

State, local, and territorial health departments move data for case notification to CDC through NNDSS. This step is voluntary, and all data is de-identified before transmission to CDC.

- CDC reports national COVID-19 case surveillance data to the World Health Organization, as required under [International Health Regulations](#). CDC also publishes COVID-19 national case surveillance data for public use at data.cdc.gov.

What data sources does CDC use for COVID-19 case surveillance?

CDC uses two data sources to obtain information on COVID-19 cases. The first is **aggregate count data**, which offer high-level information about case and death totals. Data are gathered through a robust process with the following steps:

- A CDC data team double-checks the information obtained from jurisdictions' websites via an automated overnight web review process.
- CDC additionally compares the data with information submitted by jurisdictions through a separate process.
- CDC reconciles any differences and posts the finalized information in the COVID Data Tracker.

This process is collaborative, with CDC and jurisdictions working together to ensure the accuracy of COVID-19 case and death numbers. Aggregate counts provide the most up-to-date numbers on cases and deaths. CDC may retrospectively update counts as jurisdictions provide updated information.

The second data source involves **line-level (patient-level) data**. This provides specific information for each case. CDC receives line-level data primarily from state health departments. Information is de-identified and does not include names or home addresses. CDC makes line-level data available through [patient-level data sets](#). The COVID Data Tracker also features some of this information.

Line-level data includes:

- Patient demographics such as age, race, and ethnicity
- Signs and symptoms of illness
- Underlying health conditions
- Characteristics of hospitalizations, such as ventilator use
- Clinical outcomes
- Exposures

Because it can be time-consuming for jurisdictions to collect the additional information, line-level data reporting can take more time than aggregate count reporting. CDC receives this information for most, but not all, cases.

CDC uses multiple public health surveillance systems to monitor COVID-19. This includes influenza and viral respiratory disease surveillance, [syndromic surveillance](#), lab reporting, health care systems reporting, research platforms, vital statistics, and new surveillance systems designed to answer specific questions. This combined information offers an updated picture of the spread and impact of COVID-19.

How complete are the data that CDC receives for COVID-19 cases?

The COVID-19 pandemic has strained the public health data supply chain. In many states, this has challenged hospitals, healthcare providers, and laboratories in reporting complete demographic information, such as race and ethnicity. The volume of cases has also made it challenging for state, local, and territorial health departments to conduct thorough investigations. As a result, some COVID-19 case notifications do not have complete information, even as health departments continue to make improvements through methods such as automated data flows.

Missing data can affect interpretation of factors that might put people at higher risk for severe disease. Analyses of incomplete data elements are likely an underestimate of the true occurrence.

How does CDC use COVID-19 national case surveillance data?

Case surveillance provides information on the characteristics of a disease within a population. Cases are identified using a standard [case definition](#) and are typically confirmed through laboratory testing. CDC uses national case surveillance to:

- Track the spread of COVID-19 to identify areas of concern and inform state decision-makers.
- Help state and local public health departments better control COVID-19 by evaluating trends in case demographics, exposures, and outcomes to identify groups most at risk. Examples would include healthcare workers, racial and ethnic minority groups, older adults, and people with certain underlying health conditions.
- Analyze exposure information and health outcomes among COVID-19 patients. This can help in developing guidance for the public, at-risk groups, and healthcare providers.

What are some of the challenges with national case surveillance for COVID-19?

National case surveillance data are constantly changing. As new information is gathered about previously reported cases, health departments provide updated data to CDC. As a result, surveillance data and trends from a previously reported time window might keep changing.

Another key challenge is that some people infected with the virus that causes COVID-19 have mild or no symptoms. If testing and health care services are not needed, those people are less likely to be reported as cases. Similarly, people who have had severe outcomes – such as hospitalization, intensive care unit (ICU) admission, and death – are more likely to be reported as cases. These challenges can limit analysis and interpretation of the data.

Does CDC update aggregate COVID-19 data on cases and deaths if jurisdictions submit or post the data late?

Yes, but the process varies depending on whether jurisdictions' historical data includes dates.

- If the historical data include dates, CDC incorporates it into the jurisdiction's cumulative data as soon as possible.
- If the historical data do not include dates, CDC incorporates it into the cumulative counts. This new data would be omitted from other metrics until the jurisdiction provides dates.

CDC requests that jurisdictions report date information. Here's how those scenarios are handled when specific dates cannot be provided for historical data:

- If a jurisdiction provides a date range, CDC equally distributes these data across that date range.
- If no date range is provided, CDC equally distributes the data across the first date the jurisdiction began submitting data to CDC through the date the new data were first received.

What is CDC doing to improve national case surveillance and reporting?

CDC continues to work with state, local, and territorial health departments. Goals include accelerating reporting of national case surveillance data, improving data quality, and gathering complete information about all COVID-19 cases.

Another improvement initiative involves continuing to modernize disease surveillance through [electronic case reporting](#). This process allows for automated, real-time exchange of case report information. Data flows seamlessly from a healthcare provider's electronic health record (EHR) to a public health agency. This supports timely review and action for COVID-19 cases. Electronic case reporting is a joint effort involving healthcare providers, EHR vendors, and state, local, and territorial health departments.

Where can I find information on COVID-19 cases, testing, hospitalizations, deaths, and other surveillance in the United States?

CDC offers many resources, including the following:

[CDC COVID Data Tracker](#) – This serves as CDC's home for COVID-19 data. Case and death counts reported to CDC since January 21, 2020 are available here. COVID Data Tracker is updated frequently. Timing depends on the availability of data provided by jurisdictions. Information about the frequency of updates is available on each data page. Topics covered on COVID Data Tracker include:

- COVID-19 in Your Community
- Vaccinations
- Cases, Deaths, and Testing
- Health Equity Data
- Demographic Trends
- Health Care Settings
- Genomic Surveillance
- Seroprevalence
- People at Increased Risk

[COVID Data Tracker Weekly Review](#)

key updates for the week (trends in cases, deaths, variants, laboratory testing, hospitalizations, and vaccinations)

- Interpretive summaries for trends in key COVID-19 data

Hospitalization Surveillance Network (COVID-NET)

- National hospitalization rates for COVID-19
- Characteristics of people hospitalized with COVID-19 in the U.S.

COVID-19 Serology Surveillance

- Information on large-scale geographic, community-level, and special populations seroprevalence surveys. (Results from these surveys are posted as they become available.)

COVID-19 Data from the National Center for Health Statistics (NCHS)

- Provisional death counts based on death certificate data from the National Vital Statistic System.
- Data on mental health and access to health care from the NCHS partnership with the U.S. Census Bureau on the Household Pulse Survey. (Includes indicators of anxiety and depression based on reported frequency of symptoms during the last seven days.)
- Maternal and infant characteristics among women with confirmed or presumed COVID-19 during pregnancy.

Current Hospital Capacity Estimates

- Estimated percentage of inpatient beds occupied by all patients, by state.
- Estimated percentage of inpatient beds occupied by COVID-19 patients, by state.
- Estimated percentage of ICU beds occupied by all patients, by state.

Other resources:

- CDC's [National Notifiable Diseases Surveillance System](#)
- [Electronic Case Reporting for COVID-19](#)
- [CSTE Case Definition for COVID-19](#)
- [International Health Regulations \(2005\)](#)
- [Public Health Surveillance in the United States: Evolution and Challenges](#), July 2012
- [Modernizing Centers for Disease Control and Prevention Informatics Using Surveillance Data Platform Shared Services](#) [↗](#), March-April 2018
- [CDC's Vision for Public Health Surveillance in the 21st Century](#) [📄](#), July 2012
- Centers for Disease Control and Prevention (CDC). [Introduction to Public Health](#). In: [Public Health 101 Series](#). Atlanta, GA: U.S. Department of Health and Human Services, CDC; 2014.
- [CDC MMWR Novel Coronavirus Reports](#)

Understanding COVID-19 Data

What is a COVID-19 case, including a “probable case”? How is reporting of probable cases handled by jurisdictions? [↗](#)

A COVID-19 case is an individual who has been determined to have COVID-19 using a set of criteria known as a case definition. Cases can be classified as suspect, probable, or confirmed. CDC counts include probable and confirmed cases and deaths. Suspect cases and deaths are excluded.

The case classifications for COVID-19 are described in an updated [interim COVID-19 position statement and case definition](#) issued by the Council of State and Territorial Epidemiologists. A probable case or death is defined as any one of the following:

of the following.

- Meets clinical criteria AND epidemiologic linkage with no confirmatory laboratory testing performed for SARS-CoV-2
- Meets presumptive laboratory evidence
- Meets vital records criteria with no confirmatory laboratory evidence for SARS-CoV-2

Any cases and deaths classified as probable are included in CDC case counts. The same applies to any cases and deaths classified as confirmed.

Why do the number of new cases change over time and by geographic region?

The virus that causes COVID-19 [spreads](#) easily and sustainably between people. The more closely people interact and the longer that interaction, the higher the risk of COVID-19 spread. Differences in community characteristics and changes in preventive behavior can result in increases or decreases of cases. Changes in the virus (mutations) can also lead to changes in the number of cases.

Why are the death counts on the COVID Data Tracker Cases, Deaths, and Testing page different from the provisional death counts on the COVID-19 webpage?

The count on the [Cases, Deaths, and Testing page](#) includes deaths reported by state, local, and territorial health departments. Reporting frequency might vary by jurisdiction. This reflects the most up-to-date information received by CDC based on preliminary reporting.

In contrast, [provisional COVID-19 death counts](#) from the National Center for Health Statistics (NCHS) are updated with information from death certificates. This offers the most accurate death counts, but there is a reporting lag time of one to two weeks on average. Death counts are continually updated as new death certificate data are received. For these reasons, provisional COVID-19 death counts might differ from those on other published sources.

What does the mortality rate or death rate mean?

The mortality rate is the number of people who died due to COVID-19 divided by the total number of people in the population. Since this is an ongoing outbreak, the mortality rate can change daily. CDC reports COVID-19 deaths in the [COVID Data Tracker](#).

Why do some case numbers reported by state health departments, Johns Hopkins, and the World Health Organization differ from what is posted on CDC's website?

Organizations use various methods to collect and report data, which can account for some of these differences. CDC checks overall case numbers through a confirmation process with each jurisdiction. Differences between data displayed by reporting jurisdictions and CDC's website might occur due to the timing of reporting and timing of website updates.

What are the limitations in using case surveillance data to understand the epidemiology of COVID-19?

Limitations of using case surveillance data to understand the epidemiology (who, what, where, when, how) of COVID-19 include the following:

- Case surveillance data do not represent the true burden of COVID-19 in the United States. Many people infected, even if symptomatic, do not seek medical care or get tested. In these cases, data cannot be extracted from medical records. Data can also be limited if people are unavailable or unwilling to provide information.
- Most of the case reports captured by health departments are based on laboratory reports that might contain limited patient information. Because of the volume of cases, most health departments are unable to obtain additional information on every case. As a result, many case reports are missing data on patient demographics, symptoms, underlying health conditions, characteristics of hospitalizations such as ventilator use, and other factors such as travel history. Because of missing data, analyses of these data elements are likely an underestimate of the true occurrence.
- It is difficult to capture asymptomatic cases through case surveillance. People who are asymptomatic are unlikely to seek testing unless they are identified through active screening such as contact tracing. In general, investigation of symptomatic people is prioritized.

When disease volume is high and a limited number of data elements are captured on each reported case, case surveillance data can be used to assess the following:

- Population burden
- Spread
- Increases and decreases in cases in association with mitigation strategies
- Selected demographics such as age, sex, race, ethnicity, and geography

Clinical details and other characteristics about people with COVID-19 can be better assessed through special studies. CDC conducts these [special epidemiologic studies](#) to better understand risk factors, such as underlying conditions that might put people at increased risk for serious infection. CDC also conducts special studies using hospitalization and

What are the limitations in using case surveillance data to compare jurisdictions or analyze demographic trends? 

New COVID-19 cases and deaths are recorded based on data collected and reported by state, local, and territorial health departments. This information can be affected by local testing practices, laboratory capacity, and medical resources. Comparing the COVID-19 situation among jurisdictions should not be based on these rates alone.

When studying the COVID-19 situation in these jurisdictions, the rate of new cases should be assessed alongside other data. This could include the number of tests performed, the proportion of tests that are positive for SARS-CoV-2, testing policies, excess deaths, and hospital and ICU admission rates.

In addition, jurisdictions might inconsistently report demographic data, including race and ethnicity, for COVID-19 cases. Because racial and ethnic composition varies, comparisons of COVID-19 case information should consider the population of each geographic area. For these reasons, CDC's case data might not be generalizable to the entire U.S. population.

Surveillance Reports

Does CDC create surveillance reports on COVID-19? 

Yes. On February 12, 2021, we posted the first [COVID Data Tracker Weekly Review](#), which is shared every Friday. This newsletter highlights key data from CDC's COVID Data Tracker. It summarizes important trends in the pandemic and centralizes CDC data and reporting.

The COVID Data Tracker Weekly Review replaced the COVIDView report, which was produced weekly from April 3, 2020, through February 5, 2021. An [archive of COVIDView reports](#) is maintained on the CDC website.

CDC COVID Data Tracker

What is CDC COVID Data Tracker?

[CDC COVID Data Tracker](#) is CDC's home for COVID-19 data. It provides surveillance data from across the response, including hospitalizations, vaccinations, demographic information, and daily and cumulative case and death counts reported to CDC since January 21, 2020. COVID Data Tracker is updated frequently. Timing depends on the availability of data provided by jurisdictions. Information about the frequency of updates is available on each data page. Topics covered on COVID Data Tracker include:

- COVID-19 in Your Community
- Vaccinations
- Cases, Deaths, and Testing
- Health Equity Data
- Demographic Trends
- Health Care Settings
- Genomic Surveillance
- Testing and Seroprevalence
- People at Increased Risk

How often is CDC COVID Data Tracker updated?

COVID Data Tracker is updated frequently. Timing depends on the availability of data provided by jurisdictions. Information about the frequency of updates is available on each data page.

Is it possible to download data from COVID Data Tracker?

Yes, there are multiple datasets that can be downloaded directly from COVID Data Tracker. To download data from COVID Data Tracker, navigate to the data table in the tab you are viewing and click on the download icon.



How do I download case and death data over time, from January 22, 2020, onward?

To download case and death data over time, including historical data, visit the [U.S. and State Trends page](#) and click the Download Chart button. Alternatively, the data can be downloaded at data.cdc.gov.

How do I download the most current aggregate case and death data?

To download the most current aggregate case and death data, visit the [Cases, Deaths, and Testing by State page](#) on COVID Data Tracker. Expand the "Data Table" heading, and click the Download Data button.

How do I download public use line-level case data from January 1, 2020, onward? 

You can download line-level data, including patient sex, age group, hospitalization status, and race/ethnicity, county, and state of residence (where available) from three available [COVID-19 case surveillance datasets](#).

You can conduct your own analyses using the available datasets to determine the number and selected characteristics of lab-confirmed cases shared with CDC by jurisdictions through a specific date. You can download de-identified CDC case surveillance data, which includes fields for date of first positive specimen collection, case status (lab-confirmed vs. probable), and others. See the next section titled: "CDC Publicly Available Datasets."

CDC Publicly Available Datasets

Why does CDC provide patient-level datasets to the public? 

Sharing timely and accurate COVID-19 data with the public is a core activity of CDC's COVID-19 Emergency Response as well as a key priority of [CDC's Data Modernization Initiative](#), and the administration's Executive Order on Ensuring a Data-Driven Response to COVID-19 and Future High-Consequence Public Health Threats. Publicly available datasets are critical for several reasons: open government and transparency, promotion of research, and efficiency (i.e., providing the public, media, and others access to the same data with consistency and supporting information).

How many datasets are available to the public? 

CDC has three COVID-19 case surveillance datasets:

- [COVID-19 Case Surveillance Public Use Data with Geography](#): Public use, patient-level dataset with clinical and symptom data, demographics, and state and county of residence. This dataset contains 19 data elements.
- [COVID-19 Case Surveillance Public Use Data](#): Public use, patient-level dataset with demographic and clinical information, including symptoms. No geographic data is available. This dataset contains 12 data elements.
- [COVID-19 Case Surveillance Restricted Access Detailed Data](#): Restricted access, patient-level dataset with demographic and clinical data, including symptoms. Geographic data (state and county of residence) is available. Access requires a registration process and a data-use agreement. This dataset contains 32 data elements.

What privacy protections have been applied to these datasets? 

To reduce the risk that these datasets could be used to reidentify persons, CDC designed each dataset accounting for privacy and confidentiality, and conducts ongoing privacy assessments using standard methods and systematically verifies the data prior to release. Strict privacy protections, including data suppression, were applied to all three datasets. See the information included with each dataset for more information.

Do data in these datasets match the information in the CDC COVID Data Tracker – or on state health department websites? 

Although the [CDC COVID Data Tracker](#) and health department websites also report COVID-19 case surveillance data, data might not match the CDC public use datasets due to differences in timing of the creation of the datasets and differences in the timing of reporting and case notification. The three COVID-19 case surveillance datasets are updated every two weeks, and there is a reporting lag. The CDC COVID-19 Data Tracker is updated frequently. Timing depends

on the availability of data provided by jurisdictions. Information about the frequency of updates is available on each data page. When there are differences between numbers of cases reported, data reported by health departments should be considered the most up to date for the state or territory.

Last Updated July 23, 2021